



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0043985
(43) 공개일자 2020년04월28일

- (51) 국제특허분류(Int. Cl.)
G06K 9/00 (2006.01) G06K 9/46 (2006.01)
G06K 9/62 (2006.01) G06N 3/04 (2006.01)
G06T 11/60 (2006.01) G06T 7/246 (2017.01)
- (52) CPC특허분류
G06K 9/00664 (2013.01)
G06K 9/00718 (2013.01)
- (21) 출원번호 10-2020-7004307
- (22) 출원일자(국제) 2018년07월20일
심사청구일자 없음
- (85) 번역문제출일자 2020년02월13일
- (86) 국제출원번호 PCT/US2018/043128
- (87) 국제공개번호 WO 2019/040213
국제공개일자 2019년02월28일
- (30) 우선권주장
15/683,483 2017년08월22일 미국(US)

- (71) 출원인
노스롭 그루먼 시스템즈 코퍼레이션
미국 버지니아주 22042-4511 폴스 처치 페어뷰 파크 드라이브 2980
- (72) 발명자
왕, 빅터, 와이.
미국, 캘리포니아 92131, 샌 디에고, 카미니토 아르카다 10959
칼코트, 케빈, 에이.
미국, 캘리포니아 91941, 라 메사, 쿤 에비뉴 10641
- (74) 대리인
특허법인씨엔에스

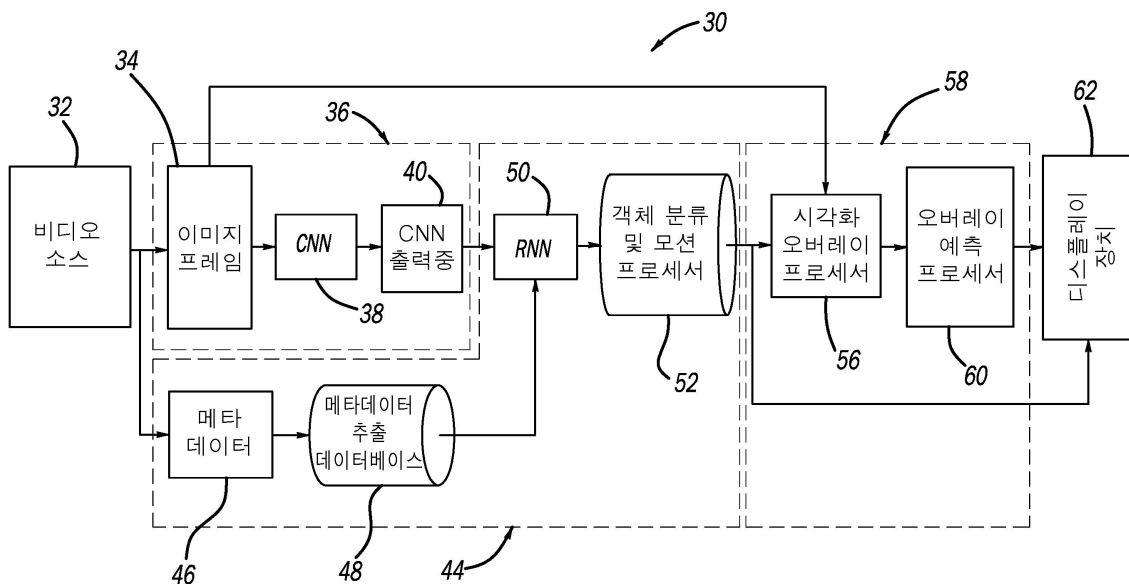
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 적응적 실시간 검출 및 검사 네트워크(ARDEN)

(57) 요약

픽셀 단위의(pixelated) 2차원 디지털 이미지의 스트림에서 객체(object)를 검출하고 인식하기 위하여 딥 러닝을 채용하는 적응적 실시간 검출 및 검사 네트워크. 네트워크는 픽셀 단위의 이미지 프레임으로서의 이미지 소스로부터의 이미지를 입력층과 출력층을 갖는 CNN에 제공하고, CNN은 이미지 내의 객체를 식별하여 분류한다. 또한, 네트워크는 이미지 소스 및 이의 위치에 관한 메타데이터(metadata)를 제공하고, 객체 분류 데이터 및 메타데이터를 이미지 내의 분류된 객체의 모션과 상대 속도를 식별하는 RNN에 제공한다. 네트워크는 CNN으로부터의 객체 분류 데이터와 RNN으로부터의 모션 데이터를 조합하고, 조합된 데이터를 상관시켜 각각의 분류된 객체 주위의 경계 상자와 분류된 객체의 상대 속도와 이동 방향의 인디케이터(indicator)를 정의하며, 이들은 디스플레이 장치에 디스플레이될 수 있다.

대표도



(52) CPC특허분류

G06K 9/4628 (2013.01)

G06K 9/6274 (2013.01)

G06K 9/6277 (2013.01)

G06N 3/0445 (2013.01)

G06N 3/0454 (2013.01)

G06T 11/60 (2013.01)

G06T 7/246 (2017.01)

G06T 2207/10016 (2013.01)

G06T 2207/20076 (2013.01)

명세서

청구범위

청구항 1

이미지 소스로부터의 비디오 스트림에서 객체(object)의 상대 속도를 식별하고, 분류하고 나타내는 방법에 있어서,

상기 비디오 스트림으로부터의 픽셀 단위의(pixelated) 이미지 프레임의 시퀀스를 입력층과 출력층을 포함하는 컨볼루션 신경망(convolutional neural network(CNN))에 제공하는 단계;

상기 CNN을 이용하여 상기 이미지 프레임 내의 객체를 식별하여 분류하고, 상기 출력층에 객체 분류 데이터를 제공하는 단계;

상기 이미지 소스로부터 메타데이터(metadata)를 제공하는 단계;

상기 출력층에서의 상기 객체 분류 데이터와 상기 메타데이터를 순환 신경망(recurrent neural network(RNN))에 제공하는 단계;

상기 RNN을 이용하여 상기 이미지 프레임 내의 분류된 상기 객체의 모션과 상대 속도를 식별하고, 상기 RNN으로부터 객체 모션 데이터를 제공하는 단계;

상기 CNN으로부터의 상기 객체 분류 데이터와 상기 RNN으로부터의 상기 객체 모션 데이터를 조합하는 단계;

조합된 상기 객체 분류 데이터와 상기 객체 모션 데이터를 상기 이미지 프레임과 상관시켜, 각각의 분류된 객체 주위의 경계 상자과 상기 분류된 객체의 상대 속도 및 이동 방향의 인디케이터(indicator)를 포함하는 상관된 이미지를 제공하는 단계; 및

상기 상관된 이미지를 디스플레이 장치 상에 디스플레이하는 단계를 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 이미지 프레임을 CNN에 제공하는 단계는, 상기 이미지 프레임을 다층 피드포워드(multi-layer feed-forward) CNN에 제공하는 단계를 포함하는, 방법.

청구항 3

제2항에 있어서,

상기 이미지 프레임을 CNN에 제공하는 단계는, 상기 이미지 프레임을 구체적으로 완전 연결층(fully connected layer)을 가지지 않는 CNN에 제공하는 단계를 포함하는, 방법.

청구항 4

제1항에 있어서,

상기 이미지 프레임을 CNN에 제공하는 단계는, 상기 이미지 프레임을 상기 CNN의 상기 입력층에 제공하는 단계를 포함하는, 방법.

청구항 5

제4항에 있어서,

상기 메타데이터를 제공하는 단계는, 상기 CNN에서의 상기 입력층에 제공된 데이터와 동일한 데이터를 포함하는 메타데이터를 제공하는 단계를 포함하는, 방법.

청구항 6

제1항에 있어서,

상기 객체 분류 데이터 및 상기 메타데이터를 RNN에 제공하는 단계는, 상기 객체 분류 데이터 및 상기 메타데이터를 LSTM(long short-term memory)을 포함하는 RNN에 제공하는 단계를 포함하는, 방법.

청구항 7

제1항에 있어서,

상기 분류된 객체의 모션 및 상대 속도를 식별하는 단계는, 방향 전환하는(turning) 객체의 장래 위치를 예측하는 단계를 포함하는, 방법.

청구항 8

제1항에 있어서,

상기 상관된 이미지 내의 상대 속도의 인디케이터는 이동하는 객체와 연관된 적어도 하나의 셰브론(chevron)이며, 상기 셰브론의 방향은 상기 객체의 이동 방향을 나타내고, 상기 셰브론의 개수는 상기 객체의 상대 속도를 나타내는, 방법.

청구항 9

제1항에 있어서,

상기 상관된 이미지 내의 상대 속도의 인디케이터는 화살표이고, 상기 화살표의 방향은 상기 객체의 이동 방향을 식별하고, 상기 화살표의 길이는 상기 객체의 상대 속도를 식별하는, 방법.

청구항 10

제1항에 있어서,

상기 비디오 소스로부터의 상기 이미지 프레임을 훈련 설비로 전송하는 단계를 더 포함하고, 상기 이미지 프레임은 상기 훈련 설비에서 훈련하는 CNN 및 훈련하는 RNN 내의 신경망 노드에서의 가중치를 훈련시키는데 사용되는, 방법.

청구항 11

제10항에 있어서,

이미지 소스로부터의 비디오 스트림에서 객체의 상대 속도를 식별하고, 분류하고 나타내는 상기 방법은, 하나의 객체 검출 및 분류 시스템에서 수행되고, 상기 방법은, 훈련된 상기 신경망 가중치를 상기 훈련 설비로부터 다른 이미지 소스로부터의 다른 비디오 스트림에서 객체의 상대 속도를 식별하고, 분류하고, 나타내는 하나 이상의 다른 객체 검출 및 분류 시스템에 전송하는 단계를 더 포함하는, 방법.

청구항 12

제11항에 있어서,

상기 다른 객체 검출 및 분류 시스템은 상기 훈련하는 CNN 및 RNN을 위하여 상기 신경망 노드에서의 상기 가중치를 더 훈련시키기 위하여 자신의 이미지 프레임을 상기 훈련 설비에 전송하는, 방법.

청구항 13

제1항에 있어서,

상기 이미지 소스는, 카메라, 전자 광학 적외선 센서, LIDAR 센서, X-선 장치, 자기 공명 촬영(magnetic resonance imaging(MRI)) 장치 및 합성 개구 레이더(synthetic aperture radar(SAR)) 장치로 이루어진 그룹으로부터 선택되는, 방법.

청구항 14

제1항에 있어서,

상기 디스플레이 장치는, 모니터, 헤드업 디스플레이(head-up display(HUD)) 장치, 고글, 프로젝터, 스마트폰 및 컴퓨터로 이루어진 그룹으로부터 선택되는, 방법.

청구항 15

이미지 소스로부터의 비디오 스트림에서 객체(object)의 상대 속도를 식별하고, 분류하고 나타내는 방법에 있어서,

상기 비디오 스트림으로부터의 픽셀 단위의(pixelated) 이미지 프레임의 시퀀스를 구체적으로 완전 연결층(fully connected layer)을 가지지 않으며 입력층과 출력층을 포함하는 컨볼루션 다층 피드포워드 신경망(convolutional multi-layer feed-forward neural network(CNN))에 제공하는 단계로서, 상기 CNN의 상기 입력층에 상기 이미지 프레임을 제공하는 단계를 포함하는 단계;

상기 CNN을 이용하여 상기 이미지 프레임 내의 객체를 식별하여 분류하고, 상기 출력층에 객체 분류 데이터를 제공하는 단계;

상기 이미지 소스로부터 메타데이터(metadata)를 제공하는 단계;

상기 출력층에서의 상기 객체 분류 데이터와 상기 메타데이터를 순환 신경망(recurrent neural network(RNN))에 제공하는 단계;

상기 RNN을 이용하여 상기 이미지 프레임 내의 분류된 상기 객체의 모션과 상대 속도를 식별하고, 상기 RNN으로부터 객체 모션 데이터를 제공하는 단계로서, 방향 전환하는(turning) 객체의 장래 위치를 예측하는 단계를 포함하는 단계;

상기 CNN으로부터의 상기 객체 분류 데이터와 상기 RNN으로부터의 상기 객체 모션 데이터를 조합하는 단계;

조합된 상기 객체 분류 데이터와 상기 객체 모션 데이터를 상기 이미지 프레임과 상관시켜, 각각의 분류된 객체 주위의 경계 상자와 상기 분류된 객체의 상대 속도 및 이동 방향의 인디케이터(indicator)를 포함하는 상관된 이미지를 제공하는 단계; 및

상기 상관된 이미지를 디스플레이 장치 상에 디스플레이하는 단계를 포함하는, 방법.

청구항 16

제15항에 있어서,

상기 메타데이터를 제공하는 단계는, 상기 CNN에서의 상기 입력층에 제공된 데이터와 동일한 데이터를 포함하는 메타데이터를 제공하는 단계를 포함하는, 방법.

청구항 17

제15항에 있어서,

상기 객체 분류 데이터 및 상기 메타데이터를 RNN에 제공하는 단계는, 상기 객체 분류 데이터 및 상기 메타데이터를 LSTM(long short-term memory)를 포함하는 RNN에 제공하는 단계를 포함하는, 방법.

청구항 18

이미지에서 객체(object)의 상대 속도를 식별하고, 분류하고 나타내기 위한 객체 검출 및 분류 시스템에 있어서,

픽셀 단위의(pixelated) 이미지 프레임의 스트림 및 메타데이터(metadata)를 제공하는 비디오 소스;

입력층과 출력층을 포함하고, 상기 입력층에서 상기 이미지 프레임에 응답하고, 상기 이미지 프레임 내의 객체를 식별 및 분류하여 상기 출력층에서 객체 분류 데이터를 제공하는 다층 피드포워드 컨볼루션 신경망(multi-layer feed-forward convolutional neural network(CNN))을 포함하는 분류 엔진;

순환 신경망(recurrent neural network(RNN))을 포함하는 예측 엔진 - 상기 RNN은 상기 출력층에서의 상기 객체

분류 데이터 및 상기 메타데이터에 응답하고, 상기 RNN은 상기 이미지 프레임 내의 분류된 상기 객체의 모션과 상대 속도를 식별하여 객체 모션 데이터를 제공하고, 상기 예측 엔진은 상기 CNN으로부터의 상기 객체 분류 데이터와 상기 RNN으로부터의 상기 객체 모션 데이터를 조합하는 객체 분류 및 모션 벡터 프로세서를 더 포함함 -;

상기 이미지 프레임과 상기 예측 엔진으로부터의 조합된 상기 객체 분류 데이터 및 상기 객체 모션 데이터에 응답하고, 조합된 상기 객체 분류 데이터 및 상기 객체 모션 데이터를 상기 이미지 프레임과 상관시켜 각각의 분류된 객체 주위의 경계 상자와 상기 분류된 객체의 상대 속도 및 이동 방향의 인디케이터를 포함하는 상관된 이미지를 제공하는 시각화 엔진; 및

상기 상관된 이미지를 디스플레이하는 디스플레이 장치를 포함하는, 객체 검출 및 분류 시스템.

청구항 19

제18항에 있어서,

상기 RNN은 방향 전환하는(turning) 객체의 위치를 예측하는, 객체 검출 및 분류 시스템.

청구항 20

제18항에 있어서,

상기 RNN은 LSTM(long short-term memory)을 포함하는, 객체 검출 및 분류 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 일반적으로 2차원 디지털 이미지에서 객체를 검출하고 분류하기 위한 시스템 및 방법에 관한 것으로, 더욱 상세하게는, 픽셀 단위의(pixelated) 2차원 디지털 이미지의 스트림에서 객체를 검출하여 분류하고, 이미지에서 검출된 객체의 위치를 식별하고, 이동하는 객체의 상대 속도 및 방향을 결정하기 위한 방법 및 시스템에 관한 것으로, 시스템은 객체를 검출하고 분류하기 위한 다층 피드포워드 컨볼루션 신경망(multi-layer feed-forward convolutional neural network(CNN))과 객체의 상대 속도를 결정하기 위한 순환 신경망(recurrent neural network(RNN))을 포함한다.

배경 기술

[0002] 인공 지능(AI)은 소프트웨어 애플리케이션들이 이들의 환경으로부터 학습하여 소정의 결과를 성취하기 위하여 그로부터 결정할 수 있게 하는 알고리즘을 채용하는 컴퓨터 사이언스의 일부이다. 기계 학습은 반복적인 방식으로 방대한 양의 원시(raw) 입력 데이터를 분석하여 데이터로부터 패턴을 추출하고 소프트웨어 애플리케이션이 어느 작업을 수행하도록 구체적으로 프로그래밍되지 않고서도 그 작업을 수행하도록 학습할 수 있게 함으로써 자신의 지식을 획득하는 소프트웨어 애플리케이션을 채용하는 AI의 일부이다. 딥 러닝(deep learning)은 증가하는 복잡한 개념들의 계층에서 소정의 실제 환경을 나타내는 것을 통해 더 큰 학습 성능을 제공하는 특정 유형의 기계 학습이다.

[0003] 딥 러닝은 통상적으로 각각의 연속하는 층이 이전 층의 출력을 수신하는 비선형 처리를 수행하는 여러 층의 신경망을 포함하는 소프트웨어 구조를 채용한다. 일반적으로, 층은 센서로부터 원시 데이터를 수신하는 입력층, 데이터로부터 추상적인 특징을 추출하는 다수의 은닉층(hidden layer) 및 은닉층으로부터의 특징 추출에 기초하여 소정의 사물(thing)을 식별하는 출력층을 포함한다. 신경망은 어떤 것이 정확한지에 대한 확률을 획득하기 위하여 노드에 대한 입력에 곱해지는 "가중치(weight)"를 각각 갖는 뉴런들 또는 노드들을 포함한다. 더욱 구체적으로는, 노드의 각각은 입력의 일부 부분인 그 노드에 대한 출력을 생성하기 위하여 노드에 대한 입력이 곱해지는 부동 소수점 숫자인 가중치를 가진다. 가중치는 초기에 "훈련되거나(trained)" 또는 신경망이 감독을 받는 처리 하에서 그리고 비용 함수를 최소화하는 것을 통해 알려진 데이터 세트를 분석하게 하여 네트워크가 가장 높은 확률을 갖는 정확한 출력을 획득할 수 있게 함으로써 설정된다.

[0004] 딥 러닝 신경망은 보통 이미지에서의 객체의 시각적 검출 및 분류를 위하여 이미지 특징 추출 및 변환을 제공하

기 위하여 채용되며, 이미지의 비디오 또는 스트림은 객체를 식별 및 분류하고 객체를 더 양호하게 인식하기 위하여 처리를 통해 학습하도록 네트워크에 의해 분석될 수 있다. 따라서, 이러한 종류의 네트워크에서, 시스템은 소정의 객체를 검출하기 위하여 동일한 처리 구성을 사용하고 객체를 인식하기 위하여 알고리즘이 학습된 방법에 기초하여 다르게 이를 분류한다.

[0005] 딥 러닝 알고리즘 및 네트워크는 데이터 처리 능력이 증가함에 따라 계속 개선된다. 개선의 특정 범위는 이미지의 검출 품질 및 객체가 인식되고 분류되는 속도를 증가시키는 차별성을 포함한다.

발명의 내용

[0006] 본 발명은 픽셀 단위의 2차원 디지털 이미지에서 객체를 검출하고 인식하기 위하여 딥 러닝을 채용하는 적응적 실시간 검출 및 검사 네트워크를 개시하고 설명한다. 네트워크는 픽셀 단위의(pixelated) 디지털 이미지 프레임으로서의 이미지 소스로부터의 이미지를 입력층과 출력층을 갖는 CNN에 제공하고, CNN은 이미지 내의 객체를 식별하고 분류한다. 또한, 네트워크는 이미지 소스 및 이의 위치에 관한 메타데이터(metadata)를 제공하고, 이미지 내의 분류된 객체의 모션과 상대 속도를 식별하고 이동하는 객체의 장래 위치를 예측하는 RNN에 객체 분류 데이터 및 메타데이터를 제공한다. 네트워크는 CNN으로부터의 객체 분류 데이터와 RNN으로부터의 모션 및 예측 데이터를 조합하고, 조합된 데이터를 상관시켜 디스플레이 장치 상에 디스플레이될 수 있는 분류된 객체의 각각 주위의 경계 상자 및 분류된 객체의 상대 속도 및 이동 방향에 대한 인디케이터를 정의한다.

[0007] 본 발명의 추가 특징들은 첨부된 도면과 함께 고려되는 이어지는 설명과 첨부된 도면으로부터 명백하게 될 것이다.

도면의 간단한 설명

[0008] 도 1은 이미지에 대한 도면이다;
 도 2는 도 1에 도시된 이미지에서 객체를 검출하고 분류하기 위한 시스템의 개략적인 블록도이다;
 도 3은 입력층, 은닉층 및 출력층을 포함하는 신경망에 대한 도면이다;
 도 4는 도 2에 도시된 시스템에 의해 처리되고 이미지에서 분류된 객체 주위의 경계 상자 및 라벨과 이동하는 각각의 객체와 연관된 세브론(chevron)을 포함하는 도 1에 도시된 이미지의 도면이고, 세브론의 개수는 객체의 상대 속도를 나타낸다;
 도 5는 도 2에 도시된 시스템에 의해 처리되고 이미지에서 분류된 객체 주위의 경계 상자 및 라벨과 이동하는 각각의 객체와 연관된 화살표를 포함하는 도 1에 도시된 이미지의 도면이고, 화살표의 길이는 객체의 상대 속도를 나타낸다; 그리고
 도 6은 기계 학습 및 훈련을 도 2에 도시된 시스템의 일부인 신경망에 제공하는 시스템의 개략적인 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0009] 픽셀 단위의(pixelated) 2차원 디지털 이미지의 스트림에서 객체의 검출, 분류 및 상대 속도를 제공하기 위한 딥 러닝을 채용하는 시스템 및 방법에 관한 본 발명의 실시예에 대한 다음의 논의는 본질적으로 단지 예시적인 것이며, 본 발명이나, 어떠한 방식으로든 이의 적용례 또는 용도를 한정하려고는 전혀 의도되지 않는다.

[0010] 도 1은 도로(14, 16)를 가로질러 정의되는 도시 교차로(12)를 도시하는 픽셀 단위의 2차원 디지털 이미지(10)의 도면이고, 인도(18)는 도로(14, 16)에 인접하게 제공된다. 변화하는 속도로 도로(14, 16)를 따라 이동하는 트럭, 자동차 등을 포함하는 다수의 차량(20)이 도시된다. 또한, 인도(18) 및 건물목(24)에서 걷고 있는 다수의 행인(22)이 도시된다. 본 명세서에서의 논의의 목적으로, 디지털 이미지(10)는 단일 스틸 이미지일 수 있거나, 이미지의 비디오 스트림으로부터의 한 프레임일 수 있다.

[0011] 아래에서 상세히 논의되는 바와 같이, 본 발명은 2차원 디지털 이미지에서의 객체, 예를 들어, 이미지(10)에서의 차량(20) 및 행인(22)의 검출 및 분류를 제공하는 딥 러닝 및 훈련된 신경망을 채용하는 적응적 실시간 검출 및 검사 네트워크(adaptive real-time detection and examination network(ARDEN))를 제공한다. 네트워크는, 실시간 비디오에서의 사용을 위하여, 모두 예를 들어 30 ms 미만인 매우 낮은 대기 시간(latency)으로, 디지털 이미지(10) 내에서 객체를 식별하고, 이미지(10) 내의 이의 픽셀 기반 위치를 제공하고, 객체 검출의 확률을 예측할 것이다. 또한, 네트워크는 이미지의 이전 프레임에서의 객체의 위치에 기초하여 이미지(10)에서 이동하는

객체의 상대 속도를 예측할 것이다.

[0012] 도 2는 위에서 언급된 ARDEN을 구체화하는 객체 검출 및 분류 시스템(30)의 블록도이다. 시스템(30)은 이미지(10)와 같은 픽셀 단위의 스틸 이미지 또는 픽셀 단위의 비디오 이미지의 스트림을 생성할 수 있는 임의의 적합한 장치일 수 있는 비디오 소스(32)를 포함한다. 적합한 비디오 소스의 비한정적인 예들은 카메라, 전자 광학 적외선 센서, LIDAR 센서, X-선 기계, 자기 공명 촬영(magnetic resonance imaging(MRI)) 장치, 합성 개구 레이더(synthetic aperture radar(SAR)) 장치 등을 포함한다. 비디오 소스(32)는, 초당 30 프레임과 같은 소정의 샘플 레이트의 픽셀 데이터로 정의되는, 비디오 데이터의 디지털 이미지 프레임(34) 시퀀스 또는 스틸 이미지를 제공한다. 시스템(30)은 비디오 프레임(34)을 수신하고 프레임(34) 내의 객체를 식별하여 분류하는 분류 엔진(36)을 포함한다. 각각의 이미지 프레임(34)은 분류 엔진(36) 내에 다층 피드포워드 컨볼루션 신경망(multi-layer feed-forward convolutional neural network(CNN))(38)을 포함하고, CNN(38)의 입력층은 이미지 프레임(34)으로부터 얻어진다. 당해 업계에서의 통상의 기술자에게 잘 알려진 바와 같이, CNN은 일반적인 매트릭스 곱셈 대신에 컨볼루션 수학 연산을 채용하고, 이미지와 같은 값들의 그리드를 처리하기 위한 특정 애플리케이션을 갖는 신경망이다. 더욱 구체적으로는, CNN(38)은 중첩하는 입력 영역 및 각각의 층에서 훈련되고 가중치가 주어진 노드를 갖는 다수의 수용(receptive) 필드 층을 포함하는 다층 신경망이고, CNN(38)은 구체적으로 완전 연결층(fully connected layer) 가지지 않으며, 이는 시스템(30)의 적응적 특징을 지원하고 제1 입력층의 더 유연한 크기 조절(resizing)을 허용한다. 더 낮은 이미지 해상도의 훈련 세트로부터의 훈련된 CNN 가중치 다른 더 높은 이미지 해상도에서의 간섭 또는 훈련을 위한 출발점으로서 사용될 수 있다. CNN(38) 내의 각각의 노드에는 이것이 후속 노드에 데이터를 어떻게 출력하는지를 정의하는 가중치 및 바이어스(bias)가 제공된다. CNN(38)의 출력층(40)은 이미지(10) 내의 검출되고 분류된 객체와 이미지(10) 내의 이의 해당하는 위치를 식별하는 원시 디지털 데이터를 제공한다. 아래에서 더 상세히 논의되는 바와 같이, 검출되고 분류된 객체는 경계 상자 및 라벨에 의해 이미지에서 식별된다.

[0013] 도 3은 조정 가능한 가중치(W)를 각각 갖는 복수의 노드(122)를 포함하는 신경망(신경 네트워크)(120)의 도면이고, 네트워크(120)는 본 명세서에서 논의되는 신경망을 일반적으로 나타내도록 의도된다. 신경망(120)은 이미지 프레임(34)으로부터 개별 픽셀 데이터를 수신하는 입력층(124)과, 픽셀 데이터에서 특징을 식별하는 2개의 은닉층(126, 128)과, 출력층(130)을 포함하며, 출력층(130) 내의 각각의 노드는 객체 중 하나를 식별한다.

[0014] 이미지(10) 내의 많은 객체는 이동하고 있다. 따라서, 이동하는 객체의 상대 속도, 이동하는 객체의 방향 및 이전 이미지 프레임에서의 상대 속도 및 방향에 기초하여 후속 이미지 프레임에서 이러한 이동하는 객체가 어디에 있을 것인지에 대한 예측과, 신경망 학습으로부터 얻어진 패턴을 식별하는 것이 바람직할 수 있다. 또한, 이동하는 객체의 위치의 예측은 선형 모션에서 이동하고 있지 않을 수 있는 방향 전환(turning)과 같은 이러한 객체의 위치를 예측하는 것을 포함한다.

[0015] 시스템(30)은 이러한 기능들을 제공하기 위하여 예측 엔진(44)을 포함한다. 비디오 소스(32)는 예측 엔진(44)에 의해 수신되는 상자 46로 표현되는 메타데이터(metadata)를 제공하고, 메타데이터(46)는 비디오 소스(32)의 위치, 즉 이의 GPS좌표, 하루 중의 시간, 기상 상태, 배터리 수명 등과 같은 특정 애플리케이션을 위한 다양하고 구체적인 정보를 포함하고, 메타데이터(46)는 시간에 있어서 이미지 프레임(34)에 상관된다. 메타데이터(46)는 특정 위치에 대한 객체 위치 예측 프로세스에 바람직한 메타데이터(46)를 선택하는 예측 엔진(44) 내의 메타데이터 추출 데이터베이스(48)에 제공된다. 일 실시예에서, 메타데이터 추출 데이터베이스(48)는 CNN(38)을 위한 입력층으로서 제공되는 이미지 프레임(34)으로부터 얻어지는 동일한 층을 사용한다.

[0016] 출력층(40)으로부터의 원시 객체 분류 데이터와 데이터베이스(48)로부터의 추출된 메타데이터는 객체가 이전 프레임(34)에서 어디에 있었는지에 기초하여 이미지 프레임(34) 내의 객체의 상대 속도와 방향을 결정하는 예측 엔진(44) 내의 순환 신경망(recurrent neural network(RNN))(50)으로 전송된다. 당해 업계의 통상의 기술자에게 알려진 바와 같이, 순환 신경망은, 훈련되고 가중치가 주어진 노드를 갖는 다수의 층을 또한 포함하며, 순차적인 데이터를 처리하기 위한 특정 애플리케이션을 갖고, 노드 사이의 연결이 방향성 사이클(directed cycle)을 형성하는 일종의 인공 신경망이다. RNN(50)의 구성은 이것이 동적인 시간적 거동을 발휘할 수 있게 하는 네트워크의 내부 상태를 형성하고, 피드포워드(feed-forward) 신경망과는 다르게, RNN은 임의의 입력 시퀀스를 처리하기 위하여 자신의 내부 메모리를 사용할 수 있다. 일 실시예에서, RNN(50)은 당해 기술 분야에서의 통상의 기술자에게 알려진 LSTM(long short-term memory) 아키텍처를 이용하여 설계된다. RNN(50)은, 메타데이터의 사용 가능성에 따라, 객체 분류, 각각의 분류된 객체의 각도(방향)의 예측, 선형 방식으로 이동하고 있지 않을 수 있는 객체의 위치의 예측, 시간의 미터법 단위 당 픽셀과 같은 입력층에만 유의미한 단위를 가질 각각의 분류된 객체의 상대 속도의 예측 및 시간의 미터법 단위 당 픽셀의 GPS 좌표 또는 다른 좌표계 기반의 이동 단위로의

변환을 포함하는 메타데이터를 형성하기 위하여, 메타데이터 추출 데이터베이스(48)와 CNN 출력층(40)로부터의 콘텐츠의 융합을 제공한다.

- [0017] 설명된 바와 같이, RNN(50)은 이것이 현재 이미지 프레임(34)과 조합하의 이전 이미지 프레임(34)을 사용하여 분류된 객체의 상대 속도를 제공할 수 있게 하는 중간 메모리 상태를 포함한다는 것이 주목된다. 시스템(30)이 단지 소스(32)로부터의 단일 스틸 이미지를 처리하고 있다면, 이전 이미지 프레임은 사용 가능하지 않고, 따라서, RNN(50)은 객체의 상대 속도의 표시를 제공할 수 없다.
- [0018] 그 다음, 분류 엔진(36)에 의해 식별된 이미지 프레임(34) 내의 객체의 모션을 식별하는 RNN(50)에 의해 생성된 데이터는 CNN(38)에 의해 제공된 이미지 프레임(34) 내의 객체의 검출을 RNN(50)에 의해 결정된 이러한 객체의 모션과 조합하는 예측 엔진(44) 내의 객체 분류 및 모션 벡터 메타데이터 프로세서(52)로 전송된다. 그 다음, 객체 위치와 객체 상대 속도를 포함하는 데이터의 조합이 이미지 프레임(34)과 함께 프레임(34) 내의 각각의 검출된 객체 주위의 경계 상자(72)와 객체 상대 속도 및 예측된 장애 위치의 표시를 제공하는 시각화 엔진(58) 내의 객체 분류 및 모션 벡터 메타데이터 프로세서(56)의 시각적 오버레이로 제공된다. 이미지(10) 내의 객체의 위치 및 상대 속도를 나타내는 이미지를 생성하는 시각화 엔진(58) 내의 이미지 프레임 및 분류 오버레이 및 모션 벡터 객체 예측 프로세서(60)에 상관 관계가 제공된다. 프로세서(52)로부터의 객체 위치 및 객체 상대 속도를 포함하는 데이터의 조합 및 이미지는 이미지를 디스플레이하기 위하여 디스플레이 장치(62)로 제공된다. 디스플레이 장치(62)는, 모니터, 헤드업 디스플레이(head-up display(HUD)) 장치, 고글, 프로젝터, 스마트폰, 컴퓨터 등과 같은, 본 명세서에 설명된 목적에 적합한 임의의 디스플레이 장치일 수 있다.
- [0019] 도 4는, 설명된 바와 같이, 시스템(30)을 통한 이미지(10)의 처리에 의해 디스플레이 장치(62) 상에 디스플레이 될 수 있는 것의 예인 이미지(70)의 예시이다. 이미지(70)에서, 식별되고 분류된 각각의 객체는 그것이 식별되었고 분류되었다는 것을 나타내고, 이의 분류, 즉 차량(vehicle), 사람(person) 등을 식별하는 라벨(74)을 포함하는 경계 상자(72)에 의해 둘러싸인다. 움직이는 분류된 객체의 각각에 대하여, 다수의 세브론(76)이 경계 상자(72)와 관련하여 제공되고, 세브론(76)의 방향은 객체의 이동 방향을 나타내고, 세브론(76)의 개수는 객체의 상대 속도를 나타낸다.
- [0020] 도 5는 객체의 각각 주위의 경계 상자(72)와 객체를 식별하는 라벨(74)을 포함하는 시스템(30)을 통한 이미지(10)의 처리에 의해 디스플레이 장치(62) 상에 디스플레이될 수 있는 것의 예인 이미지(80)의 예시이다. 그러나, 세브론(76) 대신에, 이미지(80)는 객체의 이동 방향과 객체의 상대 속도를 나타내는 화살표(82)를 포함하고, 화살표(82)의 길이는 상대 속도를 나타낸다.
- [0021] 당해 업계에서의 통상의 기술자에 의해 잘 이해되는 바와 같이, CNN(38) 및 RNN(50)에서의 노드의 각각에 대하여 적합한 가중치를 제공하도록 본 명세서에서 논의된 목적을 위하여 신경망을 훈련시키는 것이 필요하다. 이러한 훈련은 감독되며, 통상적으로, 현장에서의 사용을 위하여 가중치가 적절히 훈련되도록, 전부 오프라인에서, 이미지 내의 객체를 초기에 식별하고 노드에 대한 초기 가중치를 제공하고, 네트워크의 출력을 평가하고 조정하는 기술자를 필요로 한다. 아래에서 상세히 논의되는 바와 같이, 본 발명은 또한 CNN(38)과 RNN(50) 내에서 훈련된 노드를 이미 포함하는 도 2에 도시된 종류의 객체 검출 및 분류 시스템이 다른 객체 검출 및 분류 시스템으로부터 이미지를 수신하는 훈련 시스템에 의해 수정되고 업데이트된 노드를 가질 수 있게 하는 시스템을 포함하고, 훈련 시스템은 대표적인 신경망을 더 훈련시키기 위하여 이러한 이미지를 사용한다.
- [0022] 도 6은 본 발명의 이러한 특징을 나타내는 신경망에서 노드의 분산(distributive) 훈련 및 가중치 분포를 제공하는 객체 검출 네트워크(90)의 개략적인 블록도이다. 네트워크(90)는 훈련 설비에 위치한 훈련 시스템(92) 및 훈련 시스템(92)으로부터 분리되어 위치되고 현장에서 사용되는 복수의 물체 검출 및 분류 시스템(94)을 포함하며, 시스템(94)은 시스템(30)과 유사하다. 시스템(94)의 개수는 2 이상의 임의의 적절한 개수라는 것이 주목된다. 훈련 시스템(92)은 적합한 알고리즘을 동작시키는 컴퓨터 시스템인 훈련 도구 세트(training tool suite)(96) 및 필요한 액세서리를 포함하며, 이 모두는 당해 업계의 통상의 기술자에게 잘 알려져 있다. 기술자는 도구 세트(96)를 사용하여 시스템(94)에 의해 분류될 이미지일 수 있는 객체를 포함하는 훈련 이미지 세트를 검사하고, 훈련 이미지에서 객체 주위에 경계 박스를 그리고 이를 분류할 것이며, 기술자는 객체를 나타내는 라벨에 의해 분류된 객체를 식별한다. 상자(98)에 의해 표현되는 라벨과 이미지는 훈련될 시스템(94)에서 사용되는 동일하거나 유사한 네트워크인 훈련 CNN 및 RNN을 포함하는 분류 및 예측 엔진(100)에 제공된다. 또한, 엔진(100)은 다른 이미지에서 객체를 더욱 정확하게 분류하게 하기 위하여 그 내의 CNN 및 RNN에서의 노드에 대한 가중치의 값을 조정하는데 사용되는 평가 및 역전파(back propagation) 특징을 포함한다. 더욱 구체적으로는, 가중치가 조정되는 것에 따라 훈련 CNN 및 RNN이 객체를 정확하게 분류하도록 가중치는 제어된 방식으로 평가되

고, 가중치는 평가 함수를 통해 추출되고, 저장되고, 전송되고, 정제될 수 있으며, 훈련하는 동안의 그 전과 동안 회귀적으로 업데이트되는 정적 데이터인 특성을 가진다. 분류 및 예측 엔진(100)은 동작이 당해 업계의 통상의 기술자에게 잘 알려지고 데이터 처리량을 증가시키기 위하여 병렬 데이터 처리를 제공하는 그래픽 처리 유닛(graphical processing unit(GPU)) 클러스터(102)로 데이터를 제공하고 그로부터 데이터를 수신한다. 평가 프로세스가 완료되고 엔진(100)이 가중치가 훈련된 대상인 객체를 식별하고 분류하는 높은 예측 가능성을 제공한다고 결정하면, 가중치는 훈련된 가중치(104)로서 엔진(100)으로부터 출력된다. 또한, 훈련 시스템(92)은 훈련된 가중치(104)를 전송하고 이미지를 수신하기 위한 트랜스시버(transceiver) 또는 네트워크 라우터(106)를 포함한다.

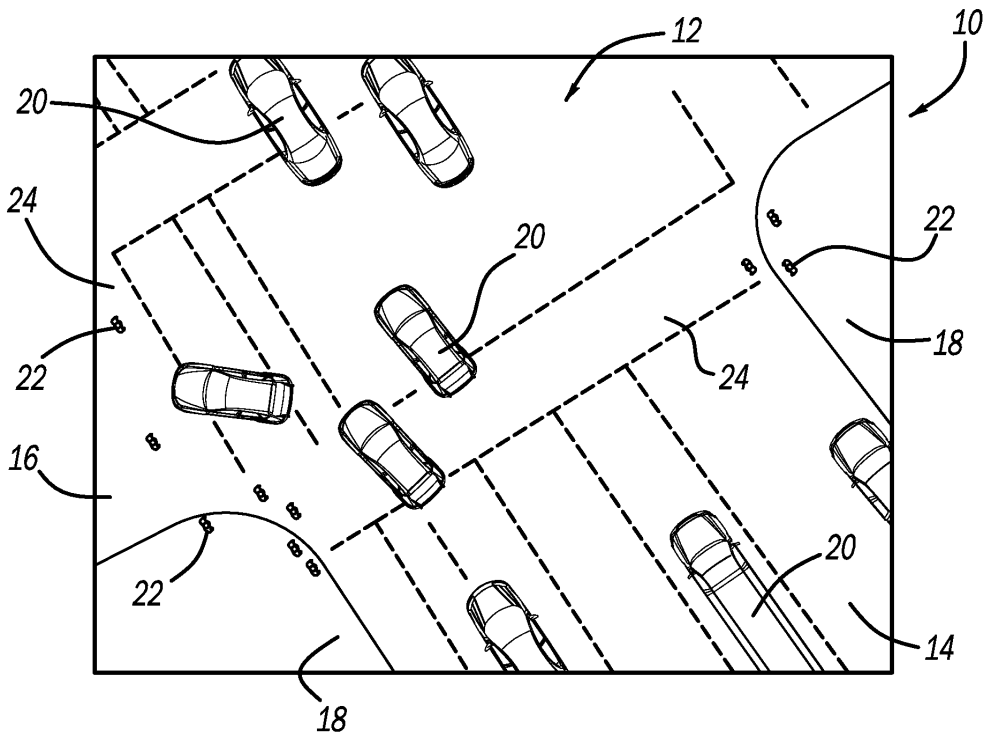
[0023] 각각의 객체 검출 및 분류 시스템(94)은 조합된 분류 엔진(36), 예측 엔진(44) 및 시각화 엔진(58)을 나타내도록 의도된 검출기(114)에 제공되는 이미지(112)를 생성하는, 비디오 소스(32)를 나타내도록 의도된, 센서(110)를 포함한다. 또한, 각각의 시스템(94)은 시스템(92) 내의 네트워크 라우터(106)와 무선 통신할 수 있게 하는 트랜스시버 또는 네트워크 라우터(116)를 포함하고, 네트워크 라우터(116)도 또한 센서(110)로부터 이미지(112)를 수신한다. 본 실시예가 서로 무선 연결된 네트워크 라우터들(106, 116)을 나타내지만, 임의의 적합한 통신 구성이 채용될 수 있다는 것이 주목된다. 이미지(112)는 네트워크 라우터(116)에 의해 훈련 시스템(92)으로 전송되고, 여기에서 이는 특정 시스템(94)의 위치에 있는 객체를 검출하는 것을 돕기 위하여 훈련 CNN 및 RNN에서의 노드를 위한 가중치의 추가 훈련을 위하여 도구 세트(96)로 제공되고 그 다음 분류 및 예측 엔진(100)으로 제공된다. 또한, 네트워크 라우터(116)는 검출기(114)에서 CNN 및 RNN에서의 가중치를 업데이트하기 위하여 상자(118)에서 검출기(114)로 제공된 훈련된 가중치(104)를 시스템(92)으로부터 수신할 수 있다. 따라서, 하나의 시스템(94)으로부터의 이미지를 이용하여 적응적으로 훈련된 가중치는 다른 하나의 시스템(94) 내의 CNN 및 RNN에서의 노드에 대하여 가중치를 업데이트하는데 사용될 수 있어, 그 시스템(94) 내의 검출기(114)가 그 시스템(94)에 의해 수신된 이미지에 있지 않을 수 있는 객체를 분류하도록 훈련될 수 있다.

[0024] 본 발명을 설명하기 위하여 본 명세서에서 논의된 여러 다양한 단계 및 과정은 전기 현상을 이용하여 데이터를 조작하고 그리고/또는 변환하는 컴퓨터, 프로세서 또는 다른 전자 계산 장치에 의해 수행되는 동작을 나타내고 있을 수 있다. 이러한 컴퓨터 및 전자 장치는 컴퓨터 또는 프로세서에 의해 수행될 수 있는 다양한 코드 또는 실행 가능한 명령어를 포함하는 실행 가능한 프로그램이 저장된 비일시적인 컴퓨터 판독 가능한 매체를 포함하는 다양한 휘발성 및/또는 비휘발성 메모리를 채용할 수 있고, 메모리 및/또는 컴퓨터 판독 가능한 매체는 모든 형태와 종류의 메모리 및 다른 컴퓨터 판독 가능한 매체를 포함할 수 있다.

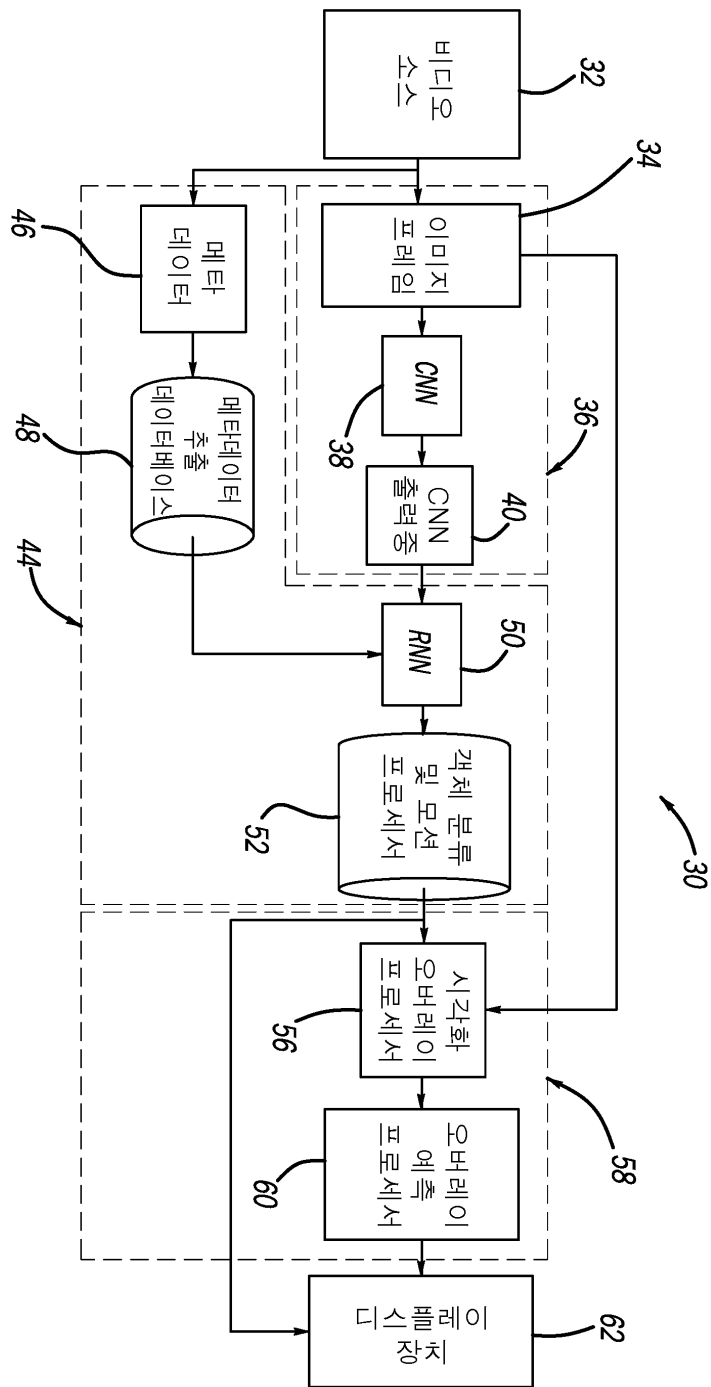
[0025] 전술한 논의는 단지 본 발명의 예시적인 실시예들을 개시하고 설명한다. 당해 업계에서의 통상의 기술자는 이러한 논의로부터 그리고 첨부된 도면과 청구범위로부터 다양한 변경, 수정 및 변동이 이어지는 청구범위에서 정의된 본 발명의 사상과 범위로부터 벗어나지 않으면서 이루어질 수 있다는 것을 쉽게 인식할 것이다.

도면

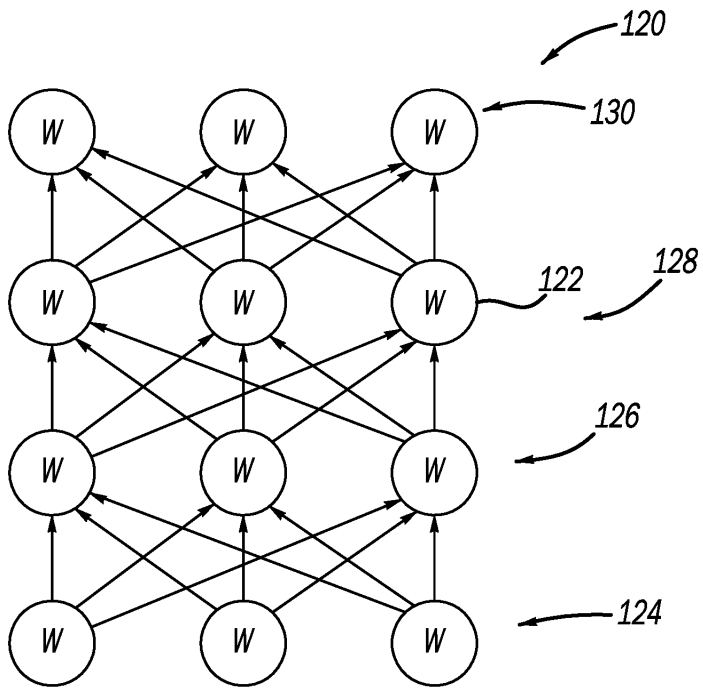
도면1



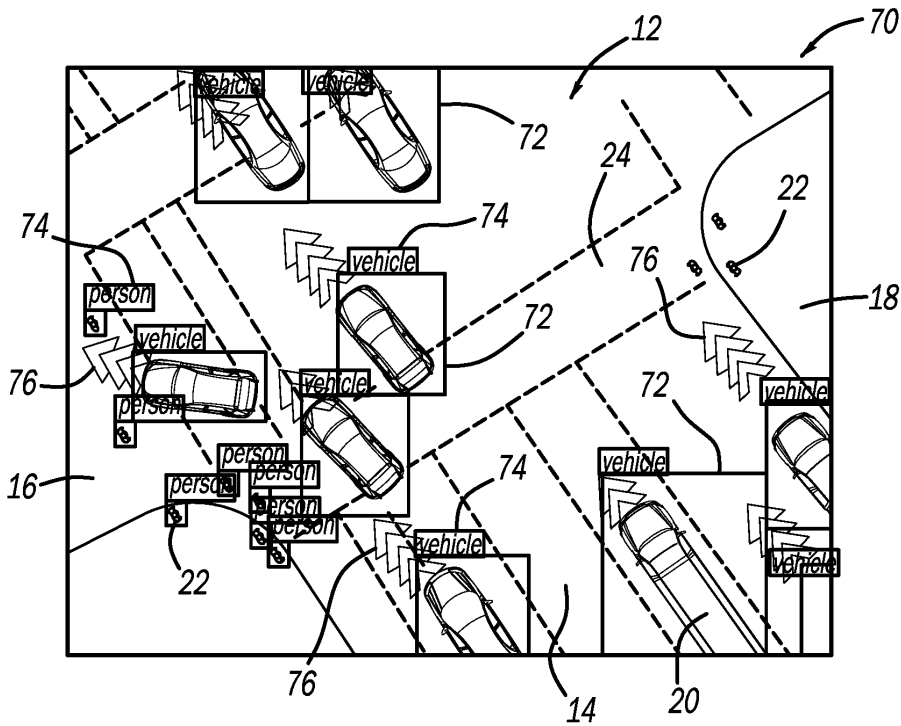
도면2



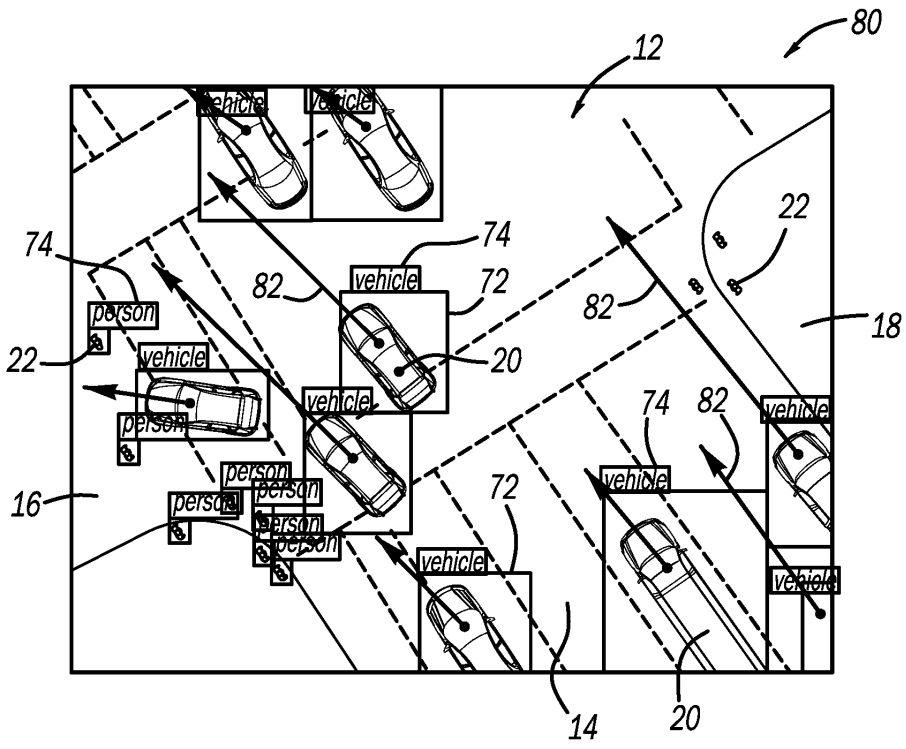
도면3



도면4



도면5



도면6

